REVIEW

Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy

G.Marius Clore and Angela M.Gronenborn

Max-Planck Institut für Biochemie, D-8033 Martinsried bei München, FRG

Introduction

Ever since the early days of biological nuclear magnetic resonance (n.m.r.) spectroscopy, it was appreciated that n.m.r. could in principle be used to determine the three-dimensional structures of proteins in solution at a resolution comparable to that afforded in the crystal state by X-ray diffraction methods. The structural information that can be extracted by n.m.r. arises from two main sources. The first and most important is the interproton nuclear Overhauser effect (NOE) which can be used to demonstrate the proximity of protons in space and to determine their separation (Noggle and Schirmer, 1971). The second is the measurement of three bond coupling constants which can be used to derive approximate torsion angle restraints from Karplus (1963) type relationships. Before such information can be extracted, however, it is essential to assign the ^{1}H-n.m.r. spectrum of the protein. This presents a formidable task and it is only in the past few years with the advent of very high field (500 MHz) n.m.r. spectrometers and new experimental techniques, in particular the whole array of two-dimensional n.m.r. experiments (Ernst et al., 1986), that this has become a feasible proposition. Once a large set of interproton distance and torsion angle restraints has been obtained, it is necessary to use computational techniques to convert this information into cartesian coordinates. This too presents a difficult problem as the n.m.r. data are limited in their number, range ($\leq$ 5 Å) and accuracy. At the present time, the structures of only a few proteins have been determined by n.m.r. and a summary of these is provided in Table I. In this article, we will present an overview of the various steps involved in solving the three-dimensional structure of a protein in solution by n.m.r. as summarized by the flow chart in Table II.

Sequential resonance assignment

The sequential resonance assignment of the ^{1}H-n.m.r. spectrum of a protein relies on two types of experiments: (i) those demonstrating through-bond connectivities, and (ii) those demonstrating through-space ($\leq$ 5 Å) connectivities. These experiments have to be performed both in H_2O and D_2O, the former to identify connectivities involving the exchangeable NH protons, and the latter to identify connectivities involving only non-exchangeable protons.

There exist a large number of two dimensional n.m.r. experiments for demonstrating homonuclear through-bond connectivities (Ernst et al., 1986). Collectively these can be called correlation experiments and are used to identify spin systems. These experiments permit one to group resonances belonging to individual amino acids, as only intraresidue connectivities are displayed, as well as to identify the spin class of amino acid to which they belong. Some spin classes have only one amino acid, so that the actual amino acid can be identified. This is the case

for Gly, Ala, Val, Thr, Leu, Ile and Lys. In other cases, a particular class may contain several amino acids. Thus Asp, Asn and Cys, and the aliphatic protons of His, Phe, Tyr and Trp, all belong to the AMX spin system (i.e. one $C^\alpha H$ proton and two $C^\beta H$ protons).

The simplest correlation experiment is the COSY experiment which was first described in 1976 by Aue et al. (1976), and demonstrates only direct though-bond connectivities. Thus, for a residue which has NH, $C^\alpha H$, $C^\beta H$ and $C^\gamma H$ protons, connectivities will only be manifested between the NH and $C^\alpha H$, $C^\alpha H$ and $C^\beta H$, and $C^\beta H$ and $C^\gamma H$ protons. The basic COSY experiment has now been superceded by more sophisticated experiments such as DQF−COSY (Rance et al., 1983) and ω1-scaled DQF−COSY (Brown, 1984) which have the advantage of exhibiting pure phase absorption diagonals when the spectra are recorded in the pure absorption mode.

Taken alone, experiments which only demonstrate direct

Table I. Summary of proteins and polypeptides whose three-dimensional structures have been determined by n.m.r.

Protein[a] (residues)	Method[b]	Reference
Micelle bound glucagon (29)	DG	Braun et al. (1983)
Insectotoxin I₅A (35)	DG	Arseniev et al. (1984)
BUSI (57)	DG	Williamson et al. (1985)
Lac repressor headpiece (51)	Model building and RD	Kaptein et al. (1985)
helix F of CRP (17)	RD	Clore et al. (1985)
α1-purothionin (45)	DG and RD	Clore et al. (1986a)
Tendasmistat (74)	RLST	Kline et al. (1986)
Metallothionein-2 (61)	RLST	Braun et al. (1986)
Phoratoxin (46)	DG and RD	Clore et al. (1987a)
Hirudin (65)	DG and RD	Clore et al. (1987b)
GH5 (79)	DG and RD	Clore et al. (1987c)
hEGF (50)	DG	Cook et al. (1987)
BPTI (56)	DG and RLST	Wagner et al. (1987)
CPI (39)	DG and RD	Clore et al. (1987d)
BSPI-2 (64)	DG and RD	Clore et al. (1987e)
Model calculations[c]		
BPTI (56)	DG	Havel and Wüthrich (1985)
BPTI (56)	RLST	Braun and Go (1986)
Crambin (46)	RD	Brünger et al. (1986) and Clore et al. (1986b)
Crambin (46)	DG	Clore et al. (1987f)

[a]The abbreviations for the various proteins are as follows: BUSI, proteinase inhibitor IIA from bull seminal plasma; CRP, cAMP receptor protein of *Escherichia coli*; GH5, globular domain of histone H5; BPTI, basic pancreatic trypsin inhibitor; CPI, potato carboxypeptidase inhibitor; hEGF, human epidermal growth factor; BSPI-2, barley serine proteinase inhibitor 2.
[b]The abbreviations for the various methods are: DG, metric matrix distance geometry; RLST, restrained least square minimization in torsion space with a variable target function; RD, restrained molecular dynamics.
[c]In the model calculations, a set of interproton distance restraints, that could realistically be determined experimentally, were derived from the X-ray structures.

Table II. Flow chart of the steps involved in determining the three-dimensional structure of a protein in solution by n.m.r.

1. *Sequential resonance assignment.*
   (i) Identification of spin systems by means of experiments demonstrating through-bond connectivities (COSY, DQF-COSY, multiple quantum spectroscopy, HOHAHA).
   (ii) Indentification of neighbouring amino acids by means of NOE measurements (NOESY) demonstrating through-space (≤ 5Å) short range ($|i-j| \leq 5$) interproton connectivities.
2. Assignment of tertiary long range ($|i-j| > 5$) NOEs.
3. Quantification or classification of NOEs to yield approximate distance restraints.
4. Measurement of three-bond coupling constants where feasible using experiments such as DQF-COSY, ω1-scaled DQF-COSY, E-COSY and z-COSY, in order to derive approximate torsion angle restraints.
5. Identification of regular secondary structure elements by means of a qualitative interpretation of the NOE data.
6. Determination of the three-dimensional structure on the basis of the approximate interproton distance and torsion angle restraints using a combination of the following approaches:
   (i) Manual and semi-automatic model building methods
   (ii) Methods relying on data bases derived from X-ray structures of proteins
   (iii) Metric matrix distance geometry
   (iv) Restrained least squares minimization in torsion angle space with variable target functions
   (v) Restrained molecular dynamics

Abbreviations used: COSY, homonuclear two-dimensional correlated spectroscopy; DQF, double quantum filtered; HOHAHA, two dimensional homonuclear Hartmann-Hahn spectroscopy; NOE, nuclear Overhauser effect; NOESY, two dimensional NOE spectroscopy.

Fig. 1. Patterns of through connectivities observed for different amino acids. In COSY spectra only direct connectivities (●) are observed, whereas in HOHAHA spectra direct (●) as well as single (○), double (□) and triple (■) relayed connectivities can be progressively observed as the mixing time in the pulse sequence is increased.

through-bond connectivities are of limited value owing to problems of spectral overlap. Thus, as one progresses from the NH and $C^\alpha H$ protons to the side chain protons, the spectral overlap tends to increase. For this reason, experiments which also demonstrate indirect or relayed through bond-connectivities, for example between NH and $C^\beta H$ protons, are invaluable. The first such experiment to be described was the relayed-COSY (Bolton and Bodenhausen, 1982), followed by homonuclear Hartmann—Hahn (HOHAHA) spectroscopy (Davis and Bax, 1985). The HOHAHA experiment is particularly versatile as one can adjust the experimental mixing time to obtain direct, single, double and multiple relayed connectivities at will. Further, the multiplet components of the cross-peaks are all in phase (i.e. have the same sign) in HOHAHA spectra, whereas in COSY type spectra they are in antiphase. This has the advantage that the HOHAHA experiment is considerably more sensitive and affords better resolution than the COSY type experiments.

Some examples of cross peak patterns that are observed in COSY and HOHAHA experiments for different amino acids are shown in Figure 1, and two examples of HOHAHA spectra recorded with different mixing times are shown in Figure 2.

Once a few spin systems have been identified, one can then proceed to identify sequential through-space connectivities involving the NH, $C^\alpha H$ and $C^\beta H$ protons by means of two dimensional NOE experiments (known as NOESY). For this purpose the most important connectivities are the $C^\alpha H(i)-NH(i + 1)$, $C^\alpha H(i)-NH(i + 2)$, $C^\alpha H(i)-NH(i + 3)$, $NH(i)-NH(i + 1)$, $C^\beta H(i)-NH(i + 1)$ and $C^\alpha H(i)-C^\beta H(i + 3)$ connectivities (Wüthrich *et al.*, 1984). Characteristic short range NOESY connectivities are illustrated in Figure 3 and examples of some NOESY spectra are shown in Figure 4.
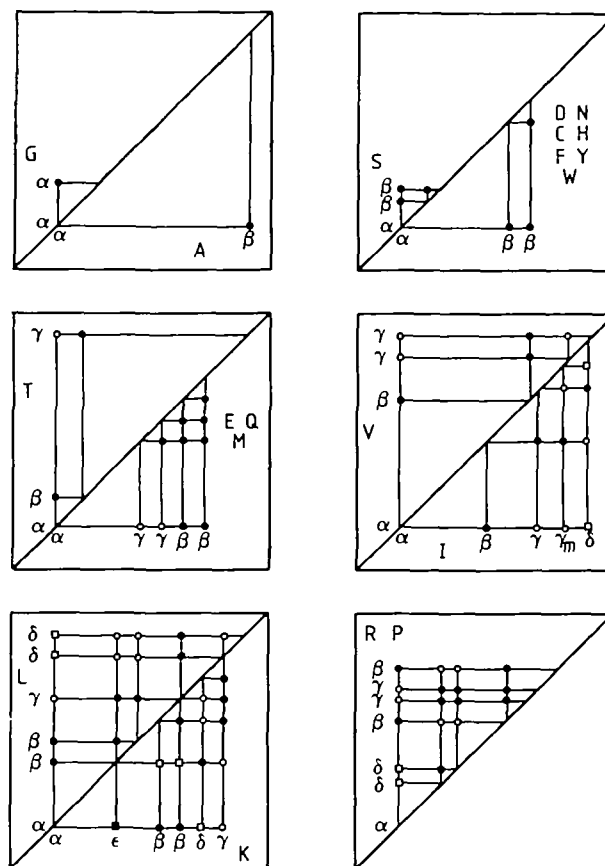
Interproton distance restraints

The initial slope of the time dependence of the NOE, N_{ij} (τ_m), between two protons i and j is equal to the cross-relaxation rate σ_{ij} between the two protons (Wagner and Wüthrich, 1979; Dobson *et al.*, 1982; Clore and Gronenborn, 1985). This is the rate at which magnetization is exchanged through space between the two protons. σ_{ij}, in turn, is proportional to $<r_{ij}^{-6}>$ and $\tau_{eff}(ij)$ (where r_{ij} is the distance between the two protons, and $\tau_{eff}(ij)$, the effective correlation time of the $i-j$ vector). It therefore follows that, at short mixing times τ_m, ratios of NOEs can yield ratios of distances, through the relationship $r_{ij}/r_{kl} \sim [N_{kl}(\tau_m)/N_{ij}(\tau_m)]^{1/6}$ providing the effective correlation times for the two interproton distance vectors are approximately the same. In the case of proteins, however, there are quite significant variations in effective correlation times as one progresses outwards from the main chain atoms to the side chain atoms, associated with higher mobility of side chains, expecially long ones. Nevertheless, because of the $<r^{-6}>$ dependence of the NOE, approximate distance restraints can be derived even in the presence of large variations in effective correlation times. Empirically, the type of classification of distances generally used is one in which strong, medium and weak NOEs correspond to distance ranges of approximately 1.8−2.7 Å, 1.8−3.3 Å and 1.8−5.0 Å, where the lower limit of 1.8 Å corresponds of the sum of the van der Waal's radii of two protons. By using such a scheme, variations in effective correlation times do *not* introduce errors
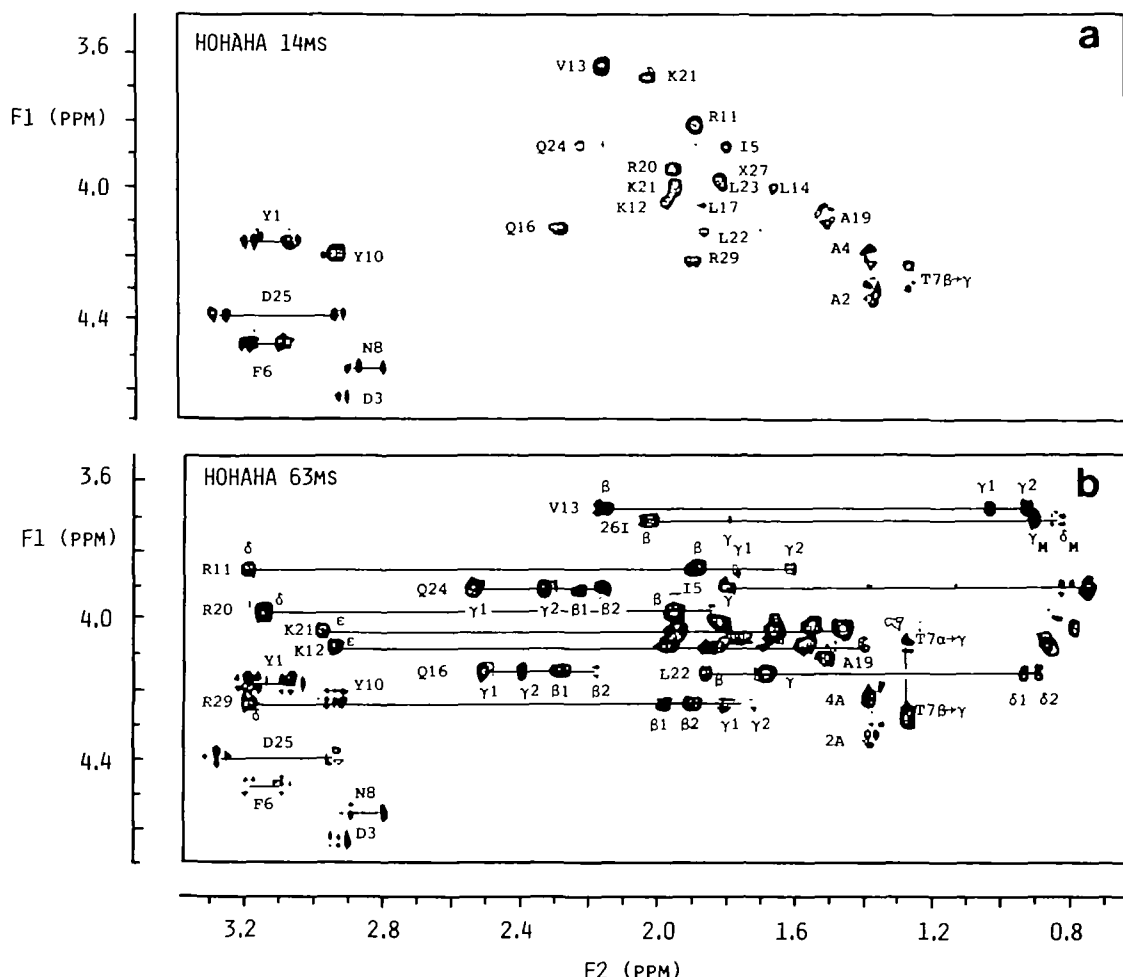
276

Fig. 2. Examples of HOHAHA spectra of the C^αH(F1 axis)-aliphatic (F2 axis) region of the (1−29) fragment of human growth hormone releasing factor showing the effect of mixing time. In (a) only direct C^αH-C^βH connecuvities are seen, while in (b) direct and multiple relayed connectivities are seen. For example, the entire spin system of the three Arg residues is apparent in (b). Residues are labelled using the one letter code and X stands for norleucine (Nle). The experimental conditions are 4 mM peptide in 20 mM phosphate buffer pH 4.0 and 30% (v/v) d₃-trifluoroethanol at 25°C (Clore et al., 1986c).
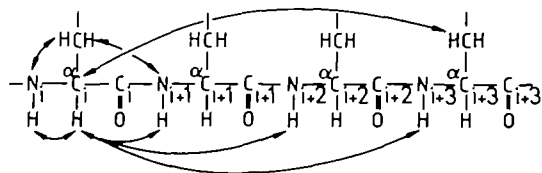
Fig. 3. Schematic diagram illustrating the sequential NOE connectivities involving the NH, C^αH and C^βH protons that can potentially be observed for a residue i in a tetrapeptide segment.

into the distance restraints. Rather they only result in an increase in the estimated range for a particular interproton distance. Thus the effect of a decrease in the effective correlation time of an interproton vector $i-j$ is simply manifested by a reduction in the magnitude of the corresponding NOE, N_{ij}, and therefore by a reduction in the precision with which that distance is defined (i.e. for a very mobile side chain, an interproton distance of say 2.3 Å is likely to appear as a weak NOE and hence be classified in the 1.8−5.0 Å range instead of the 1.8−2.7 Å range).

Torsion angle restraints

Some torsion angle restraints can be derived from three-bond coupling constants. The latter may be obtained by analysing the

multiplet patterns in COSY and COSY-like (e.g. DQF-COSY) spectra. The simplest coupling constants to determine are the $^3J_{HN\alpha}$ coupling constants which can be obtained by simply measuring the peak-to-peak separation of the antiphase components of the C^αH−NH COSY cross-peaks. Values of $^3J_{HN\alpha}$ < 6 Hz and > 9 Hz correspond to ranges of −10° to −90° and −80° to −180°, respectively, in the ϕ backbone torsion angles (Pardi et al., 1984). Considerable care, however, should be taken in deriving ϕ backbone torsion angle ranges from apparent values of $^3J_{HN\alpha}$ coupling constants measured in this way, as the minimum separation between the antiphase components of a COSY cross-peak is equal to ∼0.58 times the linewidth at half-height (Nenhaus et al., 1985).

Secondary structure

Regular secondary structure elements can easily be identified from a qualitative interpretation of the sequential NOEs as each type of secondary structure is characterized by a particular pattern of short range ($|i-j| \geq 5$) NOEs (Wüthrich et al., 1984; Wütherich, 1986; Wagner et al., 1986). Thus, for example, helices are characterized by a stretch of strong or medium NH(i)−HN(i+1), C^αH(i)−NH(i+3) and C^αH(i)−C^βH(i+3) NOEs and weak C^αH(i)−NH(i+1) NOEs. Strands, on the other
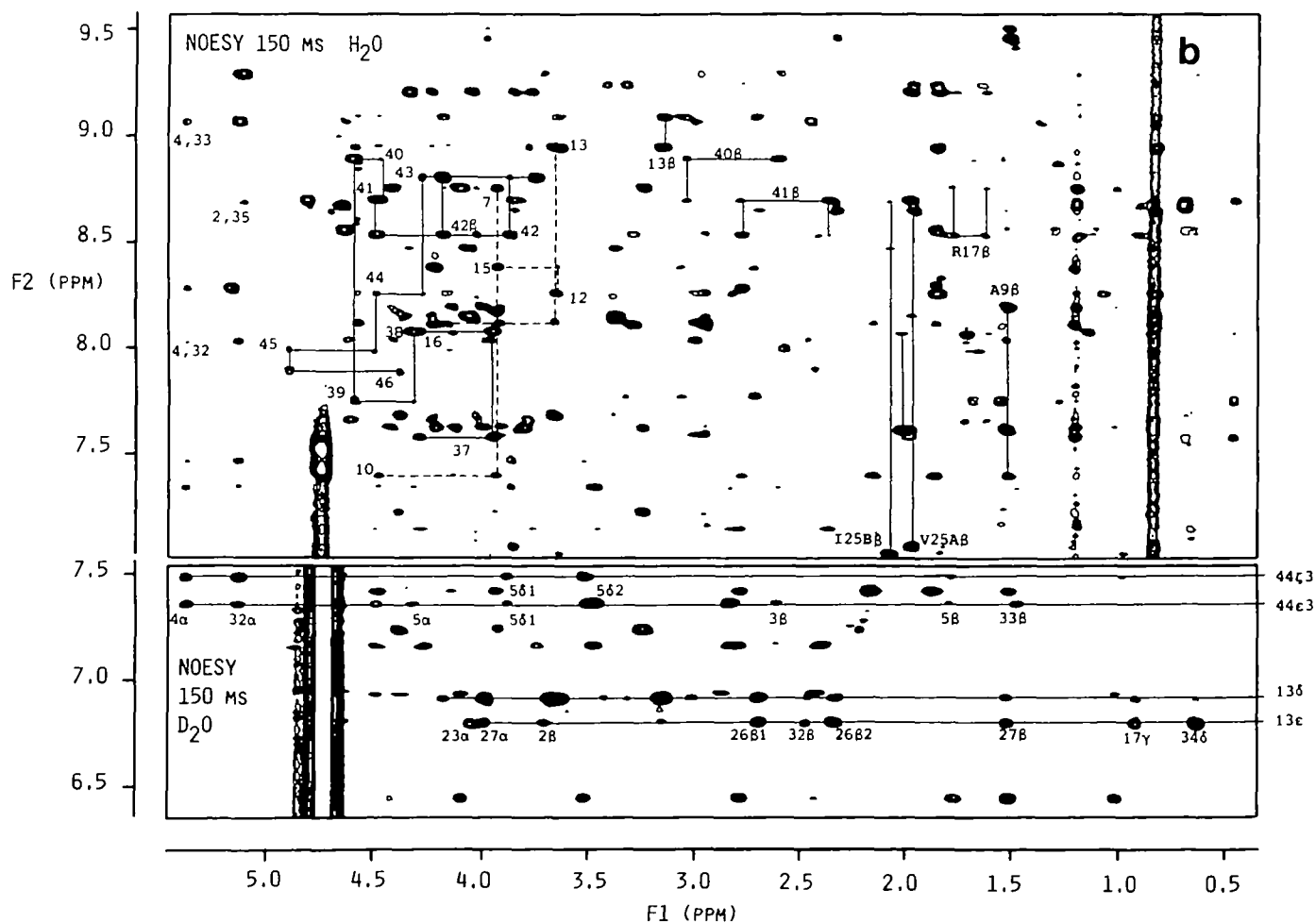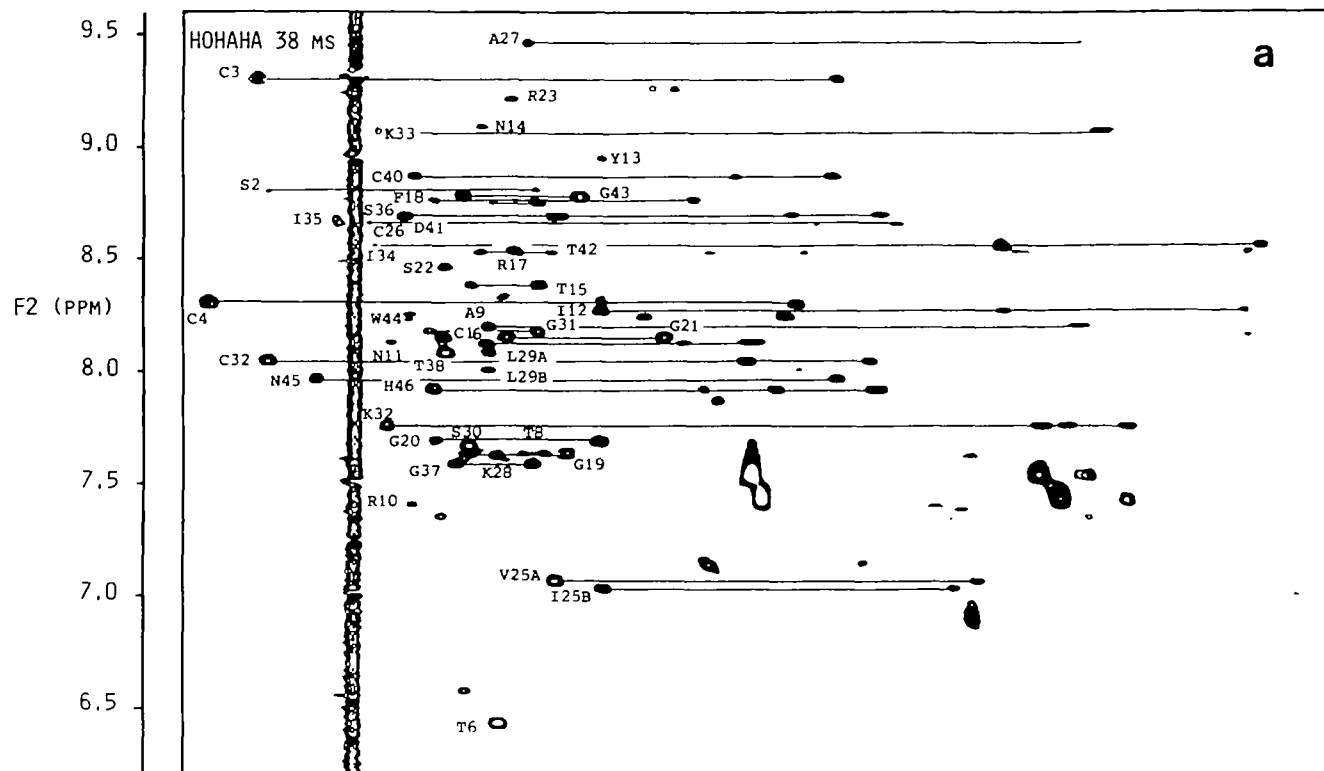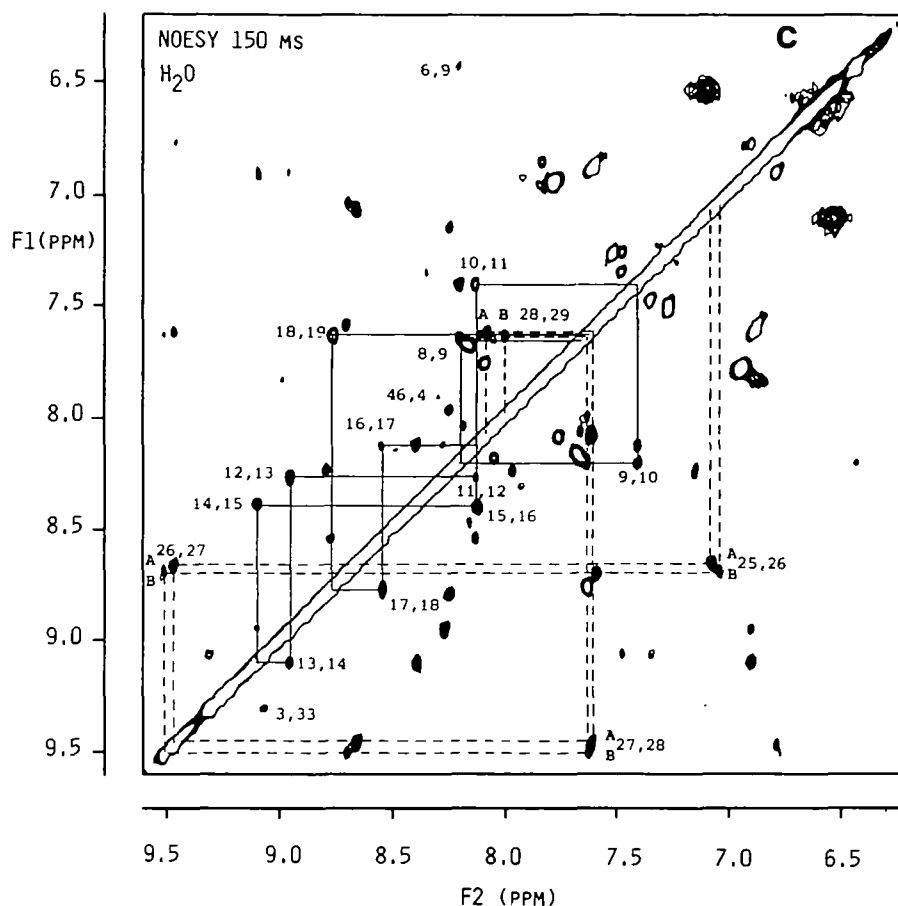
Fig. 4. HOHAHA (a) NOESY (b and c) spectra for phoratoxin (Clore *et al.*, 1987a). The NH/aromatic(F2 axis)-aliphatic(F1 axis) region is displayed in (a) and (b) while the NH/aromatic(F1 axis)-NH/aromatic(F2 axis) is displayed in (c). In the HOHAHA spectrum (a) some relayed connectivities are indicated by continuous lines and the labels are at the positions of the direct NH-C$^\alpha$H cross-peaks. The upper NOESY spectrum in (b) was recorded in H$_2$O and the lower one, showing mainly aromatic to aliphatic connectivities, in D$_2$O. Some d$_{\alpha N}$(*i,i* + 1) and d$_{\alpha N}$(*i,i* + 3) NOESY connectivities indicated by continuous (————) and dashed (———————) lines, respectively, and the peaks are labelled by residue number at the positions of the NH(*i*)−C$^\alpha$H(*i*) intraresidue NOESY cross-peaks. Also indicated in the H$_2$O spectrum are some long range NH(*i*)−C$^\alpha$H(*j*) NOEs and some d$_{\beta N}$(*i,i* + 1) connectivities with the peaks labelled by residues number followed by the letter β at the position of the NH(*i*)−C$^\beta$H(*i*) intraresidue NOESY cross-peaks. Long range interresidue NOEs from the C$^{\beta 3}$H and C$^{\epsilon 3}$H protons of Trp 44 and the C$^\delta$H and C$^\epsilon$H protons of Tyr 13 are indicated in the D$_2$O spectrum. In (c) the sequence of d$_{NN}$(*i,i* + 1) NOE connectivities from residues 8 to 19 and 25 to 29 is indicated by continuous (————) and dashed (———————) lines, respectively. There are two sets of NH peaks for residues 25−29 corresponding to phoratoxin A and B, which are present in a ratio of approximately 5:1. Experimental conditions: 8.6 mM phoratoxin in either 90% H$_2$O/10% D$_2$O or 99.96% D$_2$O at pH 3.1 and 25°C.

hand, are characterized by strong C$^\alpha$H(*i*)−NH(*i*+1) NOEs and the absence of other short range NOEs involving the NH and C$^\alpha$H protons. β-sheets can be identified and aligned from interstrand NOEs involving the NH and C$^\alpha$H protons. A summary of the expected NOEs for different types of regular secondary structure elements is given in Figure 5, and an example of the application of this approach for the small protein α1-purothionin is shown in Figure 6. It should also be pointed out that the identification of secondary structure elements can be aided by NH exchange data (i.e. slowly exchanging NH protons are usually involved in hydrogen bonds) and $^3J_{HN\alpha}$ coupling constant data.

In assessing the accuracy of the secondary structure deduced using this approach several factors should be borne in mind. Essentially, it is a data base approach in so far that the expected patterns of short range NOE connectivities for the different secondary structure elements have been derived by examining the values of all the short range distance involving the NH, C$^\alpha$H and C$^\beta$H protons in regular secondary structure elements present in protein X-ray structures. Thus, it tends to perform relatively poorly in regions of irregular structure such as loops. In

Fig. 5. Diagram of the pattern of short range NOEs involving the NH and C$^\alpha$H protons observed in different types of secondary structure. The relative intensities of the NOEs are represented by the thickness of the lines.

addition, the exact start and end of helices tend to be rather ill-defined, particularly as the pattern of NOEs for turns is not all too dissimilar from that present in helices. Thus, a turn at the end of a helix could be misinterpreted as still being part of the helix. In the case of β-sheets, the definition of the start and end is more
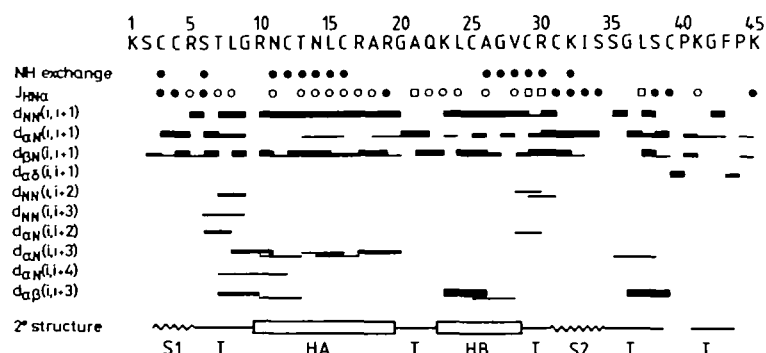
Fig. 6. Summary of the short range interresidue NOEs involving the NH, C$^{\alpha}$H and C$^{\beta}$H protons observed for α1-purothionin together with the secondary structure deduced from them (Clore *et al.*, 1987h). The relative intensities of the NOEs are indicated by the thickness of the lines. NH protons that are still present after dissolving the protein in D_2O are indicated by filled circles, and apparent values of $^3J_{HN\alpha} < 6$ Hz, 6 Hz $< ^3J_{HN\alpha} < 9$ Hz and $^3J_{HN\alpha} > 9$ Hz are indicated by the symbols O, □ and ●, respectively. The symbols for the secondary structure elements are as follows: HA and HB are helices, S1 and S2 are small strands which form a mini-antiparallel β-sheet, and T are turns.

accurate as the alignment is accomplished from the interstrand NOEs involving the NH and C$^{\alpha}$H protons.

Assignment of long range ($|i-j| > 5$) NOEs

Before the tertiary structure of a protein can be determined, it is necessary to identify as many long range (i.e. tertiary) NOEs as possible, as these are essential for determining the polypeptide fold of the protein. Once complete assignments are available, many such long range NOEs can be identified in a straightforward manner. It is usually the case, however, that the assignment of a number of long range NOE cross-peaks remains ambiguous due to resonance overlap. In some cases, such ambiguities can be readily resolved using model building on the basis of the available data (i.e. the secondary structure and the assigned long range NOEs) to derive a low resolution structure.

Tertiary structure determination

A number of different approaches can be used to determine the three-dimensional structure of a protein from n.m.r. experimental data.

The simplest approach, at least conceptually, is model building. This can be carried out either with real models or by means of interactive molecular graphics. It suffers, however, from the disadvantage that it is unable to provide an unbiased measure of the size of conformational space consistent with the n.m.r. data. Nevertheless, model building can play an important role in the early stages of a structure determination, particularly with respect to resolving ambiguities in the assignments of some of the long range NOEs.

Another approach that will no doubt become increasingly used to generate low resolution structures is based on the elegant method proposed by Kraulis and Jones (1987) which combines the NOE data with known substructures taken from the crystallographic protein data bank. The underlying philosophy of this approach is the notion that most local structural features are already well represented by existing protein structures. The method relies on only short range NOEs involving the NH, C$^{\alpha}$H and C$^{\beta}$H protons. The data base is composed of a series of distance matrices (comprising nine combinations of distances involving the NH, C$^{\alpha}$H and C$^{\beta}$H protons) generated from the protein X-ray structures in the protein data bank. A zone in the sequence of the protein under investigation is then chosen of five to eight residues in length and a set of mini distance matrices is generated from the NOE data. These mini-matrices are then

simultaneously slid along the diagonals of the precalculated distance matrices in the data base until a best fit(s) is obtained and the coordinates of the associated known fragment(s) are extracted and saved. This procedure is repeated for overlapping fragments of the sequence and the entire protein structure is then built by superimposing sequentially overlapping parts. The result is a polyalanine representation of the protein. Because the method does not make use of any long range NOEs, it can be applied at an early stage in the investigation. Long range NOEs which have been unambiguously assigned can be used to confirm the general correctness of the global fold. In addition, the structure(s) can be used as an aid to resolving ambiguities in the assignments of other long range NOEs.

The other methods used to generate structures from n.m.r. data are not based on a data base approach. There are three of these. They all have very large radii of convergence and provide an unbiased and reliable measure of the size of the conformational space consistent with the n.m.r. data. These are metric matrix distance geometry algorithms (Kuntz *et al.*, 1976, 1979; Crippen, 1977; Crippen and Havel, 1978; Wako and Scheraga, 1982; Havel *et al.* 1983; Havel and Wütherich, 1984, 1985; Sippl and Scheraga, 1986), restrained least square minimization in torsion angle space with a variable target function (Braun and Go, 1985), and restrained molecular dynamics (Clore *et al.*, 1985, 1986b; Brünger *et al.*, 1986). Of the three, metric matrix distance geometry calculations do not require an initial structure as all the calculations are carried out in n-dimensional distance space. In the case of the latter two methods, initial structures are required. These can be (i) random structures, (ii) structures that are very far from the final structure (e.g. a completely extended strand), or (iii) structures generated by the metric matrix distance geometry calculations. They should not, however, comprise structures derived by model building as this inevitably biases the final outcome. A flow chart of the calculational strategy that we have used in solving protein structures in our laboratory is shown in Figure 7. All three methods are comparable in convergence power. In general, however, the structures generated by restrained molecular dynamics are better in energetic terms than the structures generated by the other two methods, particularly with respect to the non-bonded interactions. It is our view, therefore, that all converged structures should be subjected to refinement by restrained molecular dynamics. In our experience not only does this result in large improvements in the non-bonded contacts but it also improves significantly the agreement with the experimental n.m.r. data.
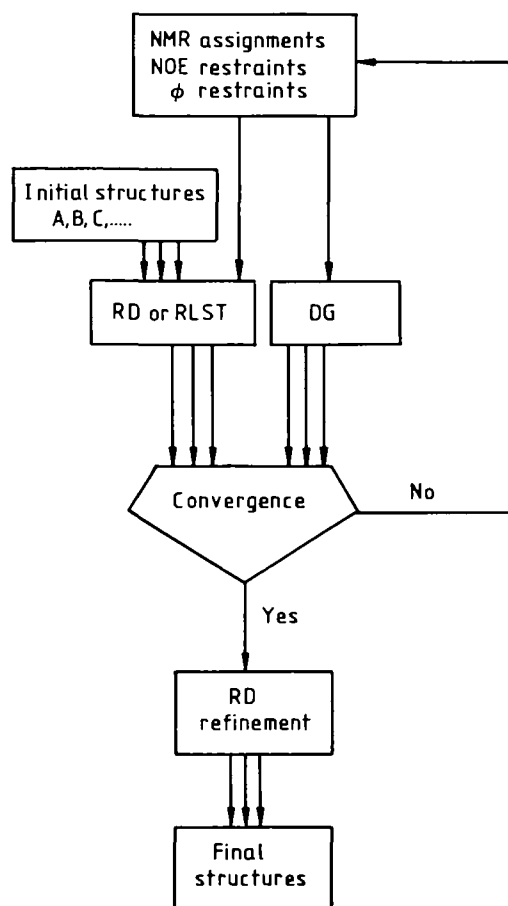
Fig. 7. Flowchart summarizing the strategy that is used in our laboratory for the determination of three-dimensional structures of proteins on the basis of n m.r. data. DG, metric matrix distance geometry (program DISGEO); RLST, restrained least squares minimization in torsion angle space with a variable target function (DISMAN); RD, restrained molecular dynamics.

The metric matrix distance geometry methods are based solely on the use of distance and planarity restraints comprising interproton distance restraints, and distances derived from torsion angle restraints, bond lengths, bond angles, planes and soft van der Waal's repulsion terms. The most widely used program to date with respect to n.m.r. applications is DISGEO (Havel, 1986). The calculations generally proceed in four phases. In the first phase a complete set of bounds on the distances between all atoms of the molecule is determined by triangulation from the NOE interproton distance restraints and the distance and planarity restraints derived from the primary structure. The latter consist of assumed exact distances between all covalently bonded and geminal pairs of atoms, as well as lower limits on the distances between all atoms more than three bonds apart (which are assumed to be no larger than the sum of the atom hard sphere radii). In the second phase a set of substructures is embedded, consistent with the bounds corresponding to distances between a subset of all the atoms. In protein applications such a suitable subset would comprise the main chain C, C^α, N and C^αH atoms as well as all non-terminal C^β and C $^\gamma$ atoms. This is followed by the third phase in which a set of initial structures which approximately fit all the data is computed. This involves choosing approximate distances at random within the triangle limits between all pairs of atoms not in the substructures, given the distances between all atoms in the substructures. This procedure is known as metrization and is the most time consuming

part of the calculation. The distances are then converted into cartesian coordinates by projection from n dimensional distance space into cartesian coordinate space, and the resulting coordinates are subjected to restrained least squares refinement with respect to all the distances in the final phase.

Like the metric matrix distance geometry methods, restrained least squares refinement in torsion space relies on distance restraints alone. The bond lengths and angles are kept fixed during the minimization and only the torsion angles are varied. What distinguishes this method from other minimization methods is the use of a variable target function such that interproton distances involving residues further and further apart in the sequence are gradually incorporated into the target function. At present, the only program implementing this method is DISMAN (Braun and Go, 1985).

Unlike the other two methods, restrained molecular dynamics also includes full energetic considerations during the entire course of the structure determination. It involves the simultaneous solution of the classical equations of motion for all atoms in the system for a suitable time period (McCammon et al., 1977, 1979; Karplus and McCammon, 1983) with the interproton distance restraints (and ϕ backbone torsion angle restraints if available) incorporated into the total energy function of the system in the form of effective potentials (Levitt, 1983; Clore et al., 1985, 1986b; Kaptein et al., 1985; Nilsson et al., 1986; Brünger et al., 1986). The power of the method lies in its ability to overcome local energy barriers and reliably locate the global minimum region, and it is similar in spirit to the simulated annealing process of non-linear optimization used in electronic circuit design and pattern recognition. There are a number of molecular dynamics programs available such as CHARMM (Brooks et al., 1983), AMBER (Weiner and Kollman, 1981), GROMOS (van Gunsteren and Berendsen, 1982), DISCOVER (Hagler, unpublished data) and XPLOR (Brünger, unpublished data). The one that is used in our laboratory is XPLOR which was originally derived from CHARMM and is especially adapted for the needs of restrained molecular dynamics.

Quality of the structures generated from n.m.r. data

There are two crucial questions regarding structures determined by n.m.r.: namely, how unique are they and how accurately have they been determined? To answer these questions it is essential to calculate a number of structures and to examine the degree of convergence. In the case of the metric matrix distance geometry calculations this is done by using different random number seeds for the computation of the distance matrices; in the case of the restrained least squares refinement in torsion space with a variable target function, by using random starting structures; finally, in the case of the restrained molecular dynamics calculations by either using random starting structures or by using a single initial structure which is very far from the final structure (such as an extended β-strand) and different random number seeds for the assignment of the initial velocities. If such a series of calculations results in a number of different structural types, all of which satisfy the experimental data within their error limits, then the information content of the experimental data can be deemed insufficient to determine the three-dimensional structure of the protein. Conversely, if convergence to a 'unique' structural set satisfying the experimental data is achieved, then one can be confident that a realistic and accurate picture of the actual solution structure has been obtained and that the region of conformational space occupied by the global energy minimum has been located. Finally, if convergence does not occur and,

in addition, the experimental restraints are not satisfied, it is likely that the restraints contain errors, for example due to the incorrect assignment of one or more long range NOESY cross-peaks.

The accuracy of the method is best judged from model calculations in which a set of interproton distances that could be realistically obtained from n.m.r. measurements has been derived from X-ray structures and then used to compute a set of structures using one of the above methods. The DISGEO program has been tested on basic pancreatic trypsin inhibitor (BPTI; Havel and Wüthrich, 1985) and crambin (Clore et al., 1987f), the DISMAN program on BPTI (Braun and Go, 1985), and the restrained molecular dynamics approach on crambin (Brünger et al. 1986; Clore et al., 1986b). In all cases the overall shape, size and folding of the polypeptide chain are reasonably well reproduced. The conformational space sampled by the methods appears to be approximately comparable, although somewhat larger for both DISMAN and restrained molecular dynamics than for DISGEO. In the case of DISGEO and DISMAN the structures tend to be slightly expanded relative to the X-ray structure, quite large deformations in the local backbone structures are apparent, and the non-bonded contacts are poor. In contrast, the structures produced by restrained molecular dynamics tend to be slightly compressed relative to the X-ray structures, the local structure tends to be better reproduced and the non-bonded contacts are good. Subjecting the structures obtained by DISGEO or DISMAN to restrained molecular dynamics refinement results in structures of the same quality as that obtained using restrained molecular dynamics from the beginning of the calculations, and indeed this may be the strategy of choice, given that the restrained molecular dynamics calculations are more time consuming than either the DISGEO or DISMAN ones.

In quantitative terms, the average atomic rms difference between the calculated structures and the corresponding X-ray structure is $1.5-2.5$ Å for the backbone atoms and $2.0-3.0$ Å for all atoms. Significant improvements, however, can be obtained by averaging the coordinates of the calculated structures, best fitted to each other. In the case of the restrained molecular dynamics crambin structures this resulted in a structure closer to the X-ray structure than any of the individual restrained molecular dynamics structures and reduced the atomic rms difference from 1.8 ± 0.3 Å to 1.0 Å for the backbone atoms and from 2.3 ± 0.3 Å to 1.6 Å for all atoms (Clore et al., 1986b).

The average structure itself has no physical significance in so far that it does not represent a structure that actually exists and further is extremely poor in energetic terms. The concept of an average structure, however, is very useful. It represents the mean structure about which the individual structures are randomly distributed, and can be characterized in terms of the rms error in its coordinates (simply given by $\sim$ rmsd/$\sqrt{n}$, where rmsd is the average atomic rms difference between the n structures and the average structure). In addition, it provides a reference point for measuring the atomic distribution of the individual structures.

A structure with physical significance that is closer to the average structure than any of the individual structures and is energetically comparable to the individual structures, can be generated from the average structure by restrained energy minimization (Clore et al., 1986b). Care, however, has to be taken when doing this as the non-bonding energy of the average structure is very high. Consequently, it is essential to increase the van der Waal's radii slowly from about a quarter of their usual values to their full values during the course of the calculations (Clore et al., 1986b). In the case of crambin, this resulted

in only a small increase in the atomic rms difference with respect to the X-ray structure (1.2 Å and 1.9 Å for the backbone atoms and all atoms, respectively, in the case of the structure derived from the restrained molecular dynamics calculations; Clore et al., 1986b). A best fit superposition of the individual restrained molecular dynamics crambin structures together with a best fit superposition of the restrained energy minimized average structure and the X-ray structure is shown in Figure 8.

A further test of the quality of the structures that can be obtained from n.m.r. interproton distance data as well as the usefulness of the concept of the average structure comes from a study in which the n.m.r. structures were used to solve the X-ray structure of crambin directly by molecular replacement (Brünger et al., 1987). It was shown that the correct solution of the translation and rotation functions of the Patterson search could only be obtained using either the restrained energy minimized average structure as a starting model or by averaging the individual rotation and translation functions obtained using the individual restrained dynamics structures as starting models. It was also shown that such starting models could be refined by conventional refinement techniques which reduced the R factor for the restrained energy minimized average structure from 0.43 at 4 Å resolution to 0.27 at 2 Å resolution without inclusion of water molecules. This compares to an R factor of 0.25 at 2 Å resolution for the published X-ray structure (Hendrickson and Teeter, 1981) when all water molecules are omitted in the computation of the calculated structure factors.

Comparison of solution and X-ray structures of globular proteins

To date there are three globular proteins whose structures have been solved both by X-ray crystallography and by n.m.r. spectroscopy and where a detailed comparison of the structures obtained by the two methods has been carried out: namely, basic pancreatic trypsin inhibitor (BPTI; Wagner et al., 1987), potato carboxypeptidase inhibitor (CPI; Clore et al., 1987d) and barley serine proteinase inhibitor 2 (BSPI-2; Clore et al., 1987e,g). The results obtained provide a yardstick to assess the extent to which differences between the 'solution' and X-ray structures arise (i) from genuine differences as reflected in differences between the experimental interproton distance restraints and a corresponding set of X-ray-derived restraints on the one hand, and (ii) from inadequacies in the input data used to determine the 'solution' structures as reflected by the limitations of the experimental data.

In all three cases there appear to be some genuine differences between the solution and X-ray structures as manifested by considerably larger interproton distance deviations between the calculated and experimental distances for the X-ray structures than for the computed 'solution' structures. In the case of CPI and BSPI-2, these differences are clearly of a minor nature in so far that most of the deviations can be corrected by restrained energy minimization of the X-ray structures with only minor atomic rms shifts (of the order of ~ 0.5 Å). Nevertheless, a few interproton distance deviations larger than 0.5 Å still remain after this procedure, and reflect more substantial differences. Thus, in the case of CPI there are two regions where the backbone atomic rms differences are significant. These areas are readily appreciated from the best fit superposition shown in Figure 9. The first region comprises the segment from residues $18-20$ which are part of a helix extending from residues $16-21$ in the 'solution' structures, whereas in the X-ray structure this region is slightly more distorted. The second region comprises the turn
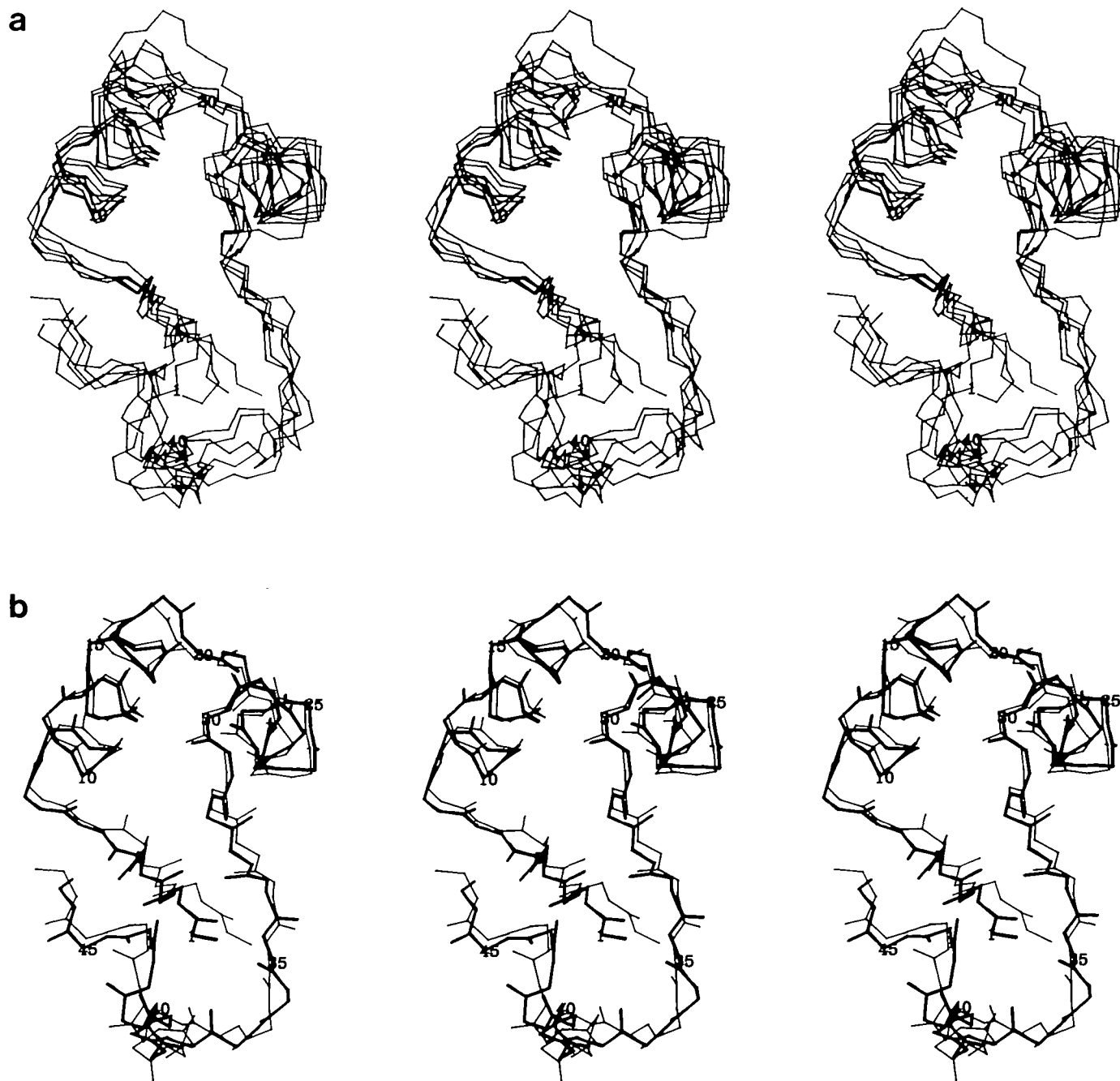
a

b

Fig. 8. Comparison of the X-ray structure of crambin with structures obtained by restrained molecular dynamics using interproton distance restraints (Clore *et al.*, 1986b). (a) Best fit superposition of the backbone (N,C$^\alpha$,C) atoms of five restrained molecular dynamics structures; (b) best fit superposition of the backbone (N,C$^\alpha$,C,O) atoms of the restrained energy minimized average structure derived from the individual restrained dynamics structures (thick lines) and the X-ray structure (thin lines). Four of the restrained dynamics structures were calculated starting from a completely extended α-strand, and the fifth structure from an extended mixed helix/strand structure (with residues 7−19 and 23−30 in the form of helices and the others in the form of extended β-strands). The calculations were based on a total of 240 approximate interproton distance restraints derived from the X-ray structure. They comprised 159 short ($|i-j| \leq 5$) and 56 long ($|i-j| > 5$) range interresidue distances and 25 intraresidue distances. All the distance restraints were ≤ 4 Å. The X-ray structure was solved by Hendrickson and Teeter (1981).

formed by residues 28−31 which has a slightly different orientation with respect to the rest of the protein in the 'solution' and X-ray structures. In addition to such differences, there are often quite large differences at the N- and C-terminal ends. These, however, cannot in general be regarded as significant as in most cases the residues at the N- and C-termini are poorly determined by the n.m.r. data (cf. Figure 9a). Rather, they simply reflect the paucity of experimental restraints at both ends of the polypeptide chain, most likely due to a larger degree of flexibility.

In quantitative terms the X-ray and computer 'solution' structures are very similar. Thus the backbone atomic rms difference between the X-ray structure and the average restrained energy minimized structure is 1.3 Å for CPI (Clore *et al.*, 1987d) and 1.5 Å for BSPI-2 (Clore *et al.*, 1987g). The corresponding values for all atoms are a little larger (2.1 Å and 2.4 Å, respectively).

In addition to the above three proteins, another protein has also been solved by n.m.r. and X-ray crystallography, namely the α-amylase inhibitor tendamistat (Kline *et al.*, 1986; Pflugrath
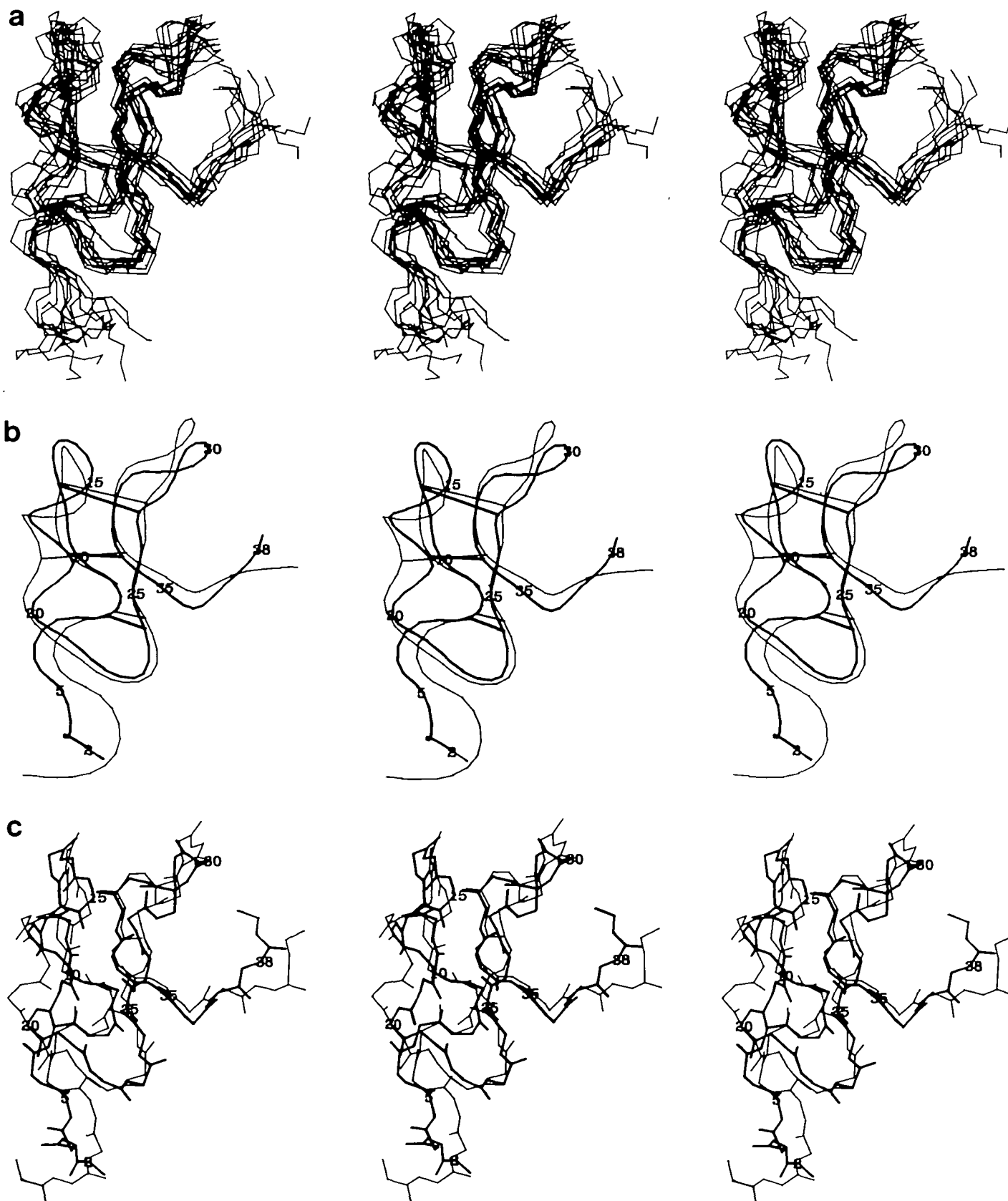
Fig. 9. Comparison of the solution and X-ray structures of potato carboxypeptidase inhibitor (Clore *et al.*, 1987d). (a) Best fit superposition of the backbone (N,C$^\alpha$,C) atoms of the eleven converged 'solution' structures; (b) and (c) best fit superposition (residues 2−38) of the restrained energy minimized average structure derived from the 11 individual structures (thick lines) and the X-ray structure (thin lines) A smoothed backbone (N,C$^\alpha$,C) atom representation with the location of the disulphide bridges indicated by lines joining the appropriate C$^\alpha$ atoms is shown in (b), and all the backbone (N,C$^\alpha$,C,O) atoms are shown in (c). The 'solution' structures were computed using a combination of metric matrix distance geometry and restrained molecular dynamics on the basis of 309 NOE restraints. The X-ray structure was solved by Rees and Lipscomb (1982).
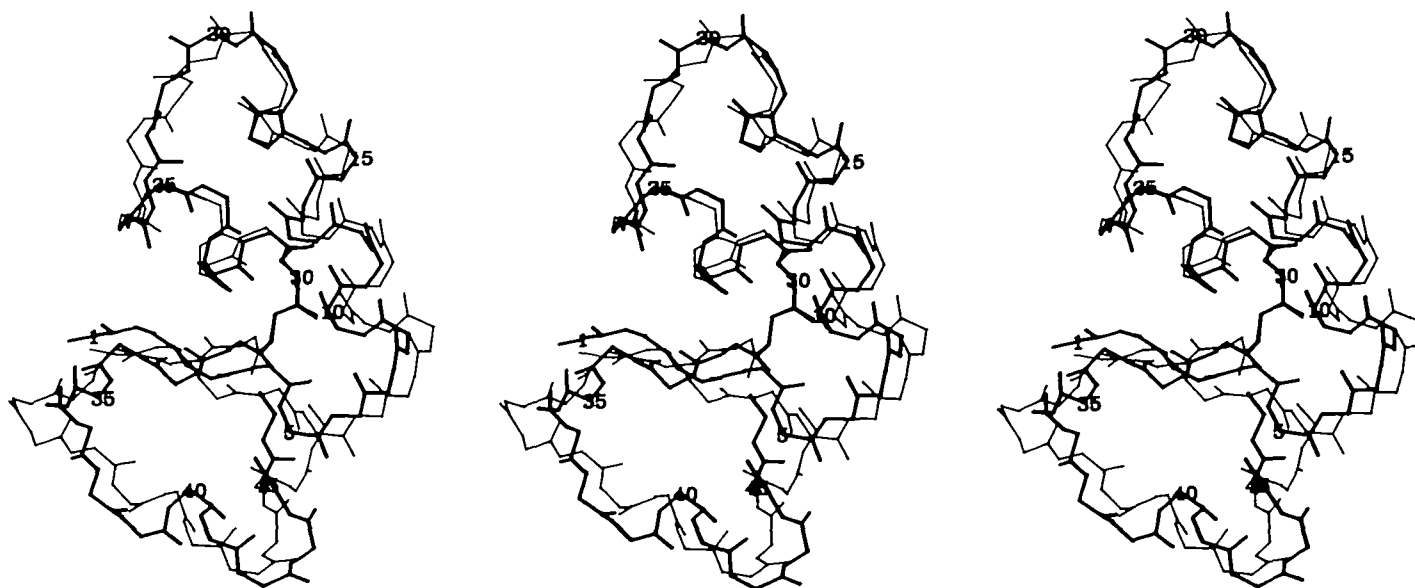
Fig. 10. Comparison of the solution structure of phoratoxin (thick lines) with the X-ray structure of crambin (thin lines) (Clore *et al.*, 1987a). The 'solution' structure shown is the restrained energy minimized average structure derived from eight independently calculated structures using a combination of metric matrix distance geometry and restrained molecular dynamics on the basis of 331 NOE restraints. All backbone (N,C$^{\alpha}$,C,O) atoms are shown. The X-ray structure of crambin was solved by Hendrickson and Teeter (1981).

et al., 1986). No detailed quantitative comparison, however, has been presented as yet. Nevertheless, it would appear that the conclusions drawn above will also apply to this protein.

Comparison of solution and X-ray structures of non-globular proteins and polypeptides

Considering non-globular proteins and polypeptides there are two cases available where the structures have been solved by n.m.r. and X-ray crystallography, namely glucagon (Braun *et al.*, 1983; Sazaki *et al.*, 1975) and metallothionein (Braun *et al.*, 1986; Furey *et al.*, 1986). In both cases there large and significant differences. Thus, the X-ray structure of glucagon is almost entirely helical, whereas the structure of micelle-bound glucagon determined by n.m.r. is composed of an irregular strand (residues 5−10) followed by two helices (residues 10−13 and 17−29) connected by a half-turn (residues 14−16). In the case of metallothionein, a two domain structure is found by both techniques and the polypeptide fold is approximately similar, but there is severe disagreement over the ligands for the seven metal ions. The coordinating amino acids for the seven Cd ions were unambiguously determined in solution by identifying scalar through-bond couplings between the Cys C$^{\beta}$H protons and the n.m.r. active (I = 1/2) ^{113}Cd^{2+} ions using heteronuclear ^{1}H-^{113}Cd COSY spectroscopy (Frey *et al.*, 1985). Thus, it is not unlikely that selective crystallization of a minor species might have occurred under the experimental conditions employed.

It would therefore appear that for non-globular proteins particular care should be taken in deriving structure−function relationships from the results of X-ray analyses alone as the structure in the crystal state may not represent the major species in solution.

Solution structures of proteins for which the X-ray structure of a related protein exists

The structure of two proteins, α1-purothionin (Clore *et al.* 1986a) and phoratoxin (Clore *et al.*, 1987a), have been solved in solution and compared with the X-ray structure of the related protein crambin (Henderickson and Teeter, 1981). The sequence

homology between these three proteins is quite high: α1-purothionin and phoratoxin exhibit sequence homologies of 33% and 39%, respectively, with respect to crambin, and 46% with respect to each other. As expected from this degree of homology, the computed solution structures of α1-purothionin and phoratoxin are close to that of the X-ray structure of crambin, as well as to each other. A best fit superposition of the phoratoxin solution and crambin X-ray structures is shown in Figure 10. Considering only the average restrained energy minimized structures, the backbone atomic rms differences with respect to crambin are 2.6 Å and 1.6 Å for α1-purothionin and phoratoxin, respectively. It is interesting to note that the α1-purothionin and phoratoxin structures are closer to crambin than to each other (backbone atomic rms differences of 3.1 Å), although their amino acid sequences are more homologous to each other than to crambin. Given that the experimental restraints used to determine the structures of α1-purothionin and phoratoxin were similar both in quality and quantity, the larger atomic rms difference between α1-purothionin and crambin may reflect the deletion of residue 24 and the presence of an extra disulphide bridge between residues 12 and 29 in α1-purothionin.

Another example is provided by the proteinase inhibitor IIA from bull seminal plasma (BUSI; Williamson *et al.*, 1985). The overall globular fold of BUSI is very close to those of the third domain of japanese quail ovomucoid (OMJPQ3; Papapmakos *et al.*, 1982) and pancreatic secretory trypsin inhibitor (PSTI; Bolognesi *et al.*, 1982), consistent with sequence homologies of 45% and 26%, respectively. A quantitative comparison was carried out for the C-terminal segment by superimposing residues 23−57 of BUSI and residues 22−56 of the two other inhibitors (the difference in alignment arising from a deletion). The C$^{\alpha}$ atomic rms difference between BUSI on the one hand and OMJPQ3 and PSTI on the other is 1.9 Å and 2.2 Å, respectively, for this segment, which compares to a C$^{\alpha}$ atomic rms difference of 1.3 Å between the two X-ray structures.

Finally, there are two other examples of solution structures where the polypeptide fold appears similar to a related X-ray structure but where no quantitative comparison has been
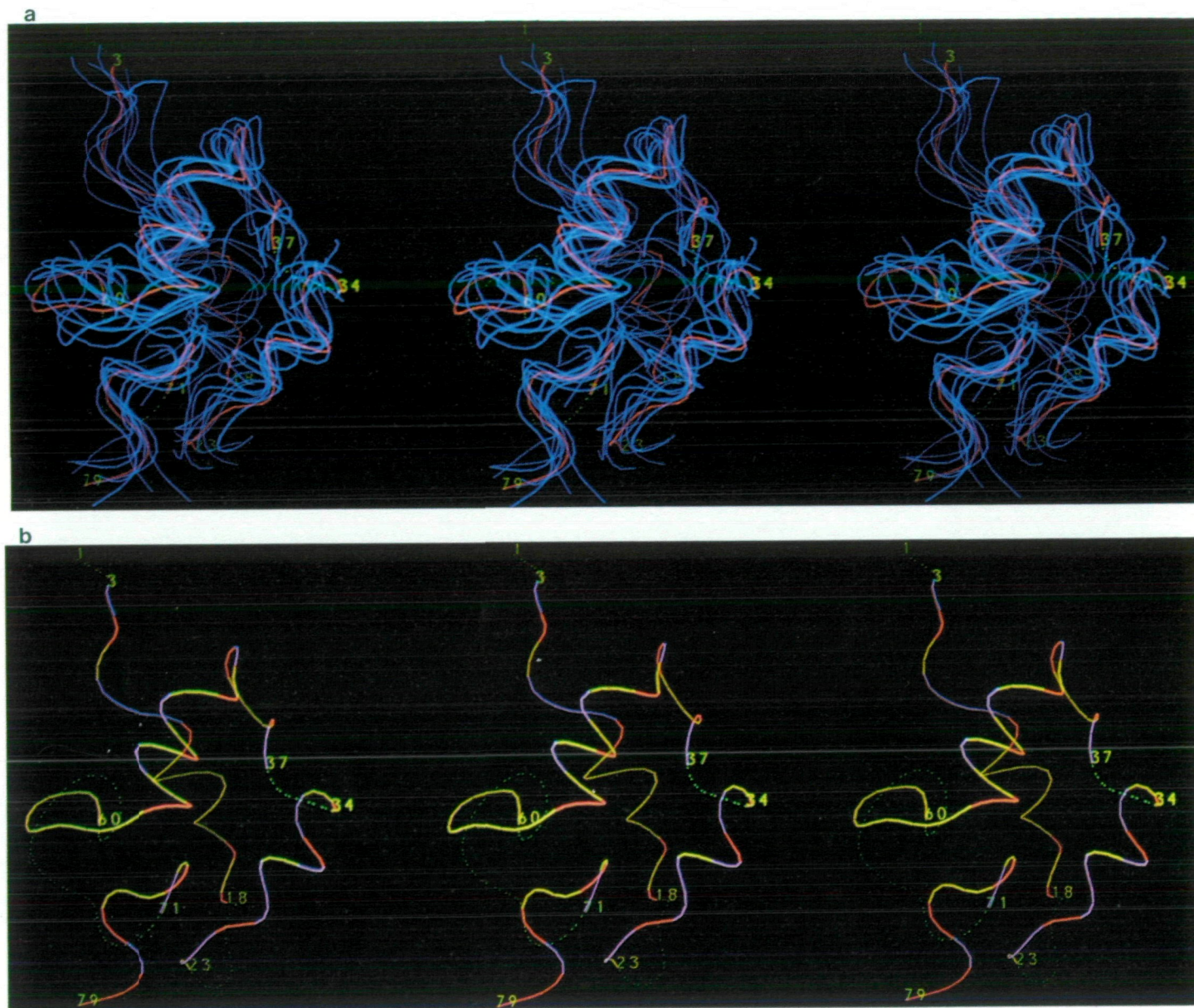
Fig. 11. Smoothed backbone (N,Cα,C) atom representation of the solution structure of the globular domain of histone H5 (Clore *et al.*, 1987c). (a) Superposition of the core residues (3−18, 22−34, 37−60 and 71−79) of the eight independently calculated structures (blue) on the restrained energy minimized average structure (red) derived from them. (b) Distribution of charged (red), polar (lilac) and hydrophobic (yellow) residues of the 'core' region of the restrained energy minimized average structure. In both (a) and (b) the two loops (residues 19−22 and 61−70) are shown as dashed lines as, although they are reasonably well defined locally, their orientation with respect to the 'core' residues could not be determined owing to the absence of any long (|*i*−*j*| >5) NOEs between the loop and 'core' residues. There are, in addition, two ill-defined regions (residues 1−2 and 35−36) which are also shown as dashed lines.

presented: namely, the 'short' scorpion insectotoxin I₅A (Arseniev *et al.*, 1984) which is similar to the helix and antiparallel β-sheet fragment in the crystal structure of the 'long' scorpion toxin v-3 (Fonticella-Camps *et al.*, 1982), and *lac* repressor headpiece (Kaptein *et al.*, 1985) which is similar to the central portion (helices 2−4) of the X-ray structure of the sequence related λCI repressor (Pabo and Lewis, 1982).

Solution structures of proteins for which there is no crystal structure

From the model calculations as well as from the results on globular proteins with known X-ray structures, it would appear that n.m.r. can be used to determine reliably the three-

dimensional structures of globular proteins within certain well defined limits. The accuracy of the coordinates are clearly nowhere near as high as those determined by X-ray crystallography (typically ≤0.3 Å at a resolution of 2.5 Å or better), and the quality of the structure depends crucially on the number and type of experimental restraints and this can only be assessed by calculating a number of structures and examining their atomic rms distribution. This is extremely important as in some cases the experimental restraints, although sufficient to determine the approximate overall polypeptide fold, may not be sufficient to determine the structure completely, such that the lack of experimental information produces variable conformations for certain parts of the protein. Two classes of ill-defined regions can occur. The ill-defined region can either be fully disordered,

or it may be locally well-defined but globally ill-defined such that the position of a whole group of residues varies from structure to structure with respect to the remainder of the protein. The structure of hirudin (Clore *et al.*, 1987b) provides a typical example of the latter case. Hirudin is a 63-residue protein which has a well defined core and two domains consisting of a finger of antiparallel β-sheet (residues $31-36$) and an exposed loop (residue $47-55$). The two minor domains are locally well defined but their relative positions with respect to the central core could not be determined as no long range (i.e. $|i-j| > 5$) NOEs could be detected between the two minor domains and the core.

Examples of other structures determined by n.m.r., for which no X-ray structure exists as yet, are the globular domain of histone H5 (Clore *et al.*, 1987c) and human epidermal growth factor (Cooke *et al.*, 1987).

Limitations of the n.m.r. method

It is clear from the above discussion that n.m.r. is a powerful method of structure determination in solution, but what are its limitations? At present it is limited to proteins of mol. wt $\leq$ 12 000. Indeed, the largest protein whose three-dimensional structure has been determined by n.m.r. is the globular domain of histone H5 (GH5) at 79 residues (Clore *et al.*, 1987c). A stereoview of the structure of GH5 is shown in Figure 11. This molecular weight limit will no doubt be raised with the introduction of ever more powerful magnets and with the development of new n.m.r. experiments, for example three-dimensional n.m.r. In addition, n.m.r. is limited to non-aggregating solutions and requires a high concentration of material (several millimolar). X-ray cystallography also has its limitations, the most obvious being the requirement of suitable crystals which diffract to high resolution and of heavy atom derivatives to solve the phase problem. In practice, therefore, it is likely that there will be relatively few proteins that will be amenable to both n.m.r. and X-ray cystallography. In such cases, the information afforded by n.m.r. and cystallography is clearly complementary, and the structures obtained by n.m.r. could potentially be used to solve the X-ray structures directly by molecular replacement.

Acknowledgements

References

Arseniev,S.A., Kondakov,V.I. Maiorov,V.N. and Bystrov,V.F. (1984) *FEBS Lett.*, 165, 65–62.

Aue,W.P., Batholdi,E. and Ernst,R.R. (1976) *J. Chem. Phys.*, 64, 2229–2246.

Bolognesi,M., Gatti,G., Menegatti,E., Guarneri,M., Marquart,M., Papamakos,E. and Huber,R. (1982) *J. Mol. Biol.*, 162, 839–868.

Bolton,P.H. and Bondenhausen,G. (1982) *Chem. Phys. Lett.*, 89, 139–144.

Braun,W. and Go,N. (1985) *J. Mol. Biol.*, 186, 611–626.

Braun,W., Wider,G., Lee,K.H. and Wüthrich,K. (1983) *J. Mol. Biol.*, 169, 921–948.

Braun,W., Wagner,G., Wörgötter,E., Vasak,M., Kagi,J.H.R. and Wüthrich,K. (1986) *J. Mol. Biol.*, 169, 921–948.

Brook,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, 4, 187–217.

Brown,L.R. (1984) *J. Magn. Reson.*, 57, 513–518.

Brünger,A.T., Clore,G.M., Gronenborn,A.M. and Karplus,M. (1986) *Proc. Natl. Acad. Sci. USA.*, 83, 3801–3805.

Brünger,A.T., Campbell,R.L., Clore,G.M., Gronenborn,A.M., Karplus,M., Petzko,G.A. and Teeter,M.M. (1987) *Science*, 235, 1049–1053.

Clore,G.M. and Gronenborn,A.M. (1985) *J. Magn. Reson.*, 61, 158–164.

Clore,G.M., Gronenborn,A.M., Brünger,A.T. and Karplus,M. (1985) *J. Mol. Biol.*, 186, 435–455.

Clore,G.M., Nilges,M., Sukumaran,D.K., Brünger,A.T., Karplus,M. and Gronenborn,A.M. (1986a) *EMBO J.*, 5, 2729–2735.

Clore,G.M., Brünger,A.T., Karplus,M. and Gronenborn,A.M. (1986b) *J. Mol. Biol.*, 191, 523–551.

Clore,G.M., Martin,S.R. and Gronenborn,A.M. (1986c) *J. Mol. Biol.*, 191, 553–561.

Clore,G.M., Sukumaran,D.K., Nilges,M. and Gronenborn,A.M. (1987a) *Biochemistry*, 26, 1732–1745.

Clore,G.M., Sukumaran,D.K., Nilges,M., Zarbock,J. and Gronenborn,A.M. (1987b) *EMBO J.*, 6, 529–537.

Clore,G.M., Gronenborn,A.M., Nilges,M. Sukumaran,D.K. and Zarbock, J. (1987c) *EMBO J.*, 6, 1833–1842.

Clore,G.M., Gronenborn,A.M., Nilges,M., and Ryan,C.A. (1987d) *Biochemistry*, in press.

Clore,G.M., Gronenborn,A.M., Kjaer,M. and Poulsen,F.M. (1987e) *Protein Engineering*, 1, 305–311.

Clore,G.M., Gronenborn,A.M., Brünger,A.T., Karplus,M. and Gronenborn,A.M. (1987f) *FEBS Lett.*, 213, 269–277.

Clore,G.M., Gronenborn,A.M., James,M.N.G., Kjaer,M., McPhalen,C.A. and Poulsen,F. (1987g) *Protein Engineering*, 1, 313–318.

Clore,G.M., Sukumaran,D.K., Gronenborn,A.M., Teeter,M.M., Whitlow,M. and Jones,B.L. (1987h) *J. Mol. Biol.*, 193, 571–578.

Cooke,R.M., Wilkinson,A.J., Baron,M., Pastore,A., Tappin,M.J., Campbell,I.D., Gregory,H. and Sheard,B. (1987) *Nature*, 327, 339–341.

Crippen,G.M. (1977) *J. Comp. Phys.*, 24, 96–107.

Crippen,G.M. and Havel,T.F. (1975) *Acta Crystallogr.*, A3A, 282–284.

Davis,D.G. and Bax,A. (1985) *J. Am. Chem. Soc.*, 107, 2821–2822.

Dobson,C.M., Olejniczak,E.T. Poulsen,F.M. and Ratcliffe,R.G. (1982) *J. Magn. Reson.*, 48, 97–110.

Ernst,R.R., Bodenhausen,G. and Wokaun,A. (1986) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Clarendon Press, Oxford.

Fontecilla-Camps,J.C., Almassy,R.J., Suddath,F.L. and Bugg,C.e. (1982) *Toxicon*, 20, 1–7.

Frey,M.H., Wagner,G., Vasak,M., Sørensen,O.W., Neuhaus,D., Kägi,J.H.R., Ernst,R.R. and Wüthrich,K. (1985) *J. Am. Chem. Soc.*, 107, 6847–6851.

Furey,W.F., Robbins,A.H., Clancy,L.L., Wunge,D.R., Wang,B.C. and Stroud,C.D. (1986) *Science*, 231, 704–709.

Havel,T.F. (1986) *DISGEO, Quantum Chemistry Program Exchange Program No. 507*, Indiana University.

Havel,T.F. and Wüthrich,K. (1984) *Bull. Math. Biol.*, 45, 673–698.

Havel,T.F. and Wüthrich,K. (1985) *J. Mol. Biol.*, 182, 281–294.

Havel,T.F., Kuntz,I.D. and Crippen,G.M. (1983) *Bull. Math. Biol.*, 45, 665–720.

Henderickson,W.A. and Teeter,M.M. (1981) *Nature*, 290, 107–112.

Kapstein,R., Zuiderweg,E.R.P., Scheek,R.M., Boelens,R. and van Gunsteren,W.F. (1985) *J. Mol. Biol.*, 182, 179–128.

Karplus,M. (1963) *J. Am. Chem. Soc.*, 85, 2870–2871.

Karplus,M. and McCammon,J.A. (1983) *Annu. Rev. Biochem.*, 52, 263–300.

Kline,A.D., Braun,W. and Wüthrich,K. (1986) *J. Mol. Biol.*, 183, 503–507.

Kraulis,P.J. and Jones,T.A. (1987) In Ehrenberg,A., Rigler,R., Gräslund,A. and Nilsson,L., (eds) *Structure, Dynamics and Function of Biomolecules.* Springer-Verlag, Berlin, pp. 118–121.

Kuntz,I.D., Crippen,G.M., Kollman,P.A. and Kimmelman,D. (1976) *J. Mol. Biol.*, 106, 983–994.

Kuntz,I.D., Crippen,G.M. and Kollman,P.A. (1979) *Biopolymers*, 18, 939–957.

Levitt,M. (1983) *J. Mol. Biol.*, 170, 723–764.

McCammon,J.A., Gelin,B.R. and Karplus,M. (1977) *Nature*, 267, 585–590.

McCammon,J.A., Wolynes,P.G. and Karplus,M. (1979) *Biochemistry*, 18, 927–942.

Neuhaus,D., Wagner,G., Vasak,M., Kägi,J.H.R. and Wüthrich,K. (1985) *Eur. J. Biochem.*, 151, 257–273.

Nilsson,L., Clore,G.M., Gronenborn,A.M., Brünger,A.T. and Karplus,M. (1986) *J. Mol. Biol.*, 188, 455–475.

Noggle,J.H. and Schirmer,R.E. (1971) *The Nuclear Overhauser Effect – Chemical Applications.* Academic Press, New York.

Pabo,C.O. and Lewis,M. (1982) *Nature*, 298, 443–447.

Papamakos,E., Weber,E., Bode,W., Huber,R., Empie,M.W., Kato,I. and Laskowski,M. (1982) *J. Mol. Biol.*, 158, 515–537.

Pardi,A., Billeter,M. and Wüthrich,K. (1984) *J. Mol. Biol.*, 180, 741–751.

Pflugrath,J., Wiegand,E., Huber,R. and Vertesy,L. (1986) *J. Mol. Biol.*, 189, 383–386.

Rance,M., Sørenson,O.W., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, 117, 479–485.

Rees,D.C. and Lipscomb,W.N. (1982) *J. Mol. Biol.*, 160, 475–498.

Sazaki,K., Dockevill,S., Ackmiak,D.A., Tickle,I.J. and Blundell,T.L. (1975) *Nature*, **257**, 751–757.

Sippl,M.J. and Scheraga,H.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2283–2287.

van Gunsteren,W.F. and Berendsen,H.J.C. (1982) *Biochem. Soc. Trans.*, **10**, 301–305.

Wagner,G. and Wüthrich,K. (1979) *J. Magn. Reson.*, **33**, 675–680.

Wagner,G. Neuhaus,D., Wörgötter,E., Vasak,M., Kagi,J.R.H. and Wüthrich,K. (1986) *J. Mol. Biol.*, **187**, 131–135.

Wagner,G., Braun,W., Havel,T.F., Schaumann,T., Go,N and Wüthrich,K. (1987) *J. Mol. Biol.*, in press.

Wako,H. and Scheraga,H.A. (1982) *J. Prot. Chem.*, **1**, 85–117.

Weiner,P.K. and Kollman,P.A. (1981) *J. Comput. Chem.*, **2**, 287–303.

Williamson,M.P., Havel,T.F. and Wüthrich,K. (1985) *J. Mol. Biol.*, **182**, 295–315.

Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids.* J. Wiley, New York.

Wüthrich,K. Billeter,M. and Braun,W. (1984) *J. Mol. Biol.*, **180**, 715–740.